

7-2013

Spatially explicit modeling of Schistosomiasis risk in eastern China based on a synthesis of epidemiological, environmental, and intermediate host genetic data

Matthias Schrader

Justus Liebig University, Giessen, Germany

Torsten Hauffe

Justus Liebig University, Giessen, Germany

Zhijie Zhang

Fudan University, Shanghai, China

George M. Davis

George Washington University

Fred Jopp

Justus Liebig University, Giessen, Germany

Follow this and additional works at: http://hsrc.himmelfarb.gwu.edu/smhs_microbio_facpubs



Part of the [Medical Immunology Commons](#), [Medical Microbiology Commons](#), [Parasitic Diseases Commons](#), and the [Parasitology Commons](#)

Recommended Citation

Schrader, M., Hauffe, T., Zhang, Z., Davis, G.M., Jopp, F., Remais, J.V., Wilke, T. (2013). Spatially explicit modeling of Schistosomiasis risk in eastern China based on a synthesis of epidemiological, environmental, and intermediate host genetic data. *PLoS Neglected Tropical Diseases*, 7(7):e2327.

This Journal Article is brought to you for free and open access by the Microbiology, Immunology, and Tropical Medicine at Health Sciences Research Commons. It has been accepted for inclusion in Microbiology, Immunology, and Tropical Medicine Faculty Publications by an authorized administrator of Health Sciences Research Commons. For more information, please contact hsrc@gwu.edu.

Authors

Matthias Schrader, Torsten Hauffe, Zhijie Zhang, George M. Davis, Fred Jopp, Justin V. Remais, and Thomas Wilke

Spatially Explicit Modeling of Schistosomiasis Risk in Eastern China Based on a Synthesis of Epidemiological, Environmental and Intermediate Host Genetic Data

Matthias Schrader¹, Torsten Hauffe¹, Zhijie Zhang², George M. Davis³, Fred Jopp¹, Justin V. Remais⁴, Thomas Wilke^{1*}

1 Department of Animal Ecology and Systematics, Justus Liebig University Giessen, Giessen, Germany, **2** Department of Epidemiology and Biostatistics, School of Public Health, Fudan University, Shanghai, People's Republic of China, **3** Department of Microbiology and Tropical Medicine, George Washington University Medical Center, Washington, District of Columbia, United States of America, **4** Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, Georgia, United States of America

Abstract

Schistosomiasis japonica is a major parasitic disease threatening millions of people in China. Though overall prevalence was greatly reduced during the second half of the past century, continued persistence in some areas and cases of re-emergence in others remain major concerns. As many regions in China are approaching disease elimination, obtaining quantitative data on *Schistosoma japonicum* parasites is increasingly difficult. This study examines the distribution of schistosomiasis in eastern China, taking advantage of the fact that the single intermediate host serves as a major transmission bottleneck. Epidemiological, population-genetic and high-resolution ecological data are combined to construct a predictive model capable of estimating the probability that schistosomiasis occurs in a target area ("spatially explicit schistosomiasis risk"). Results show that intermediate host genetic parameters are correlated with the distribution of endemic disease areas, and that five explanatory variables—altitude, minimum temperature, annual precipitation, genetic distance, and haplotype diversity—discriminate between endemic and non-endemic zones. Model predictions are correlated with human infection rates observed at the county level. Visualization of the model indicates that the highest risks of disease occur in the Dongting and Poyang lake regions, as expected, as well as in some floodplain areas of the Yangtze River. High risk areas are interconnected, suggesting the complex hydrological interplay of Dongting and Poyang lakes with the Yangtze River may be important for maintaining schistosomiasis in eastern China. Results demonstrate the value of genetic parameters for risk modeling, and particularly for reducing model prediction error. The findings have important consequences both for understanding the determinants of the current distribution of *S. japonicum* infections, and for designing future schistosomiasis surveillance and control strategies. The results also highlight how genetic information on taxa that constitute bottlenecks to disease transmission can be of value for risk modeling.

Citation: Schrader M, Hauffe T, Zhang Z, Davis GM, Jopp F, et al. (2013) Spatially Explicit Modeling of Schistosomiasis Risk in Eastern China Based on a Synthesis of Epidemiological, Environmental and Intermediate Host Genetic Data. PLoS Negl Trop Dis 7(7): e2327. doi:10.1371/journal.pntd.0002327

Editor: Darren J. Gray, University of Queensland, Australia

Received: January 14, 2013; **Accepted:** June 12, 2013; **Published:** July 25, 2013

Copyright: © 2013 Schrader et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported in part by the US NIH/NSF Ecology of Infectious Disease Program (<http://www.fic.nih.gov>) through grant no. 0622743 (JVR) and by the NIAID Tropical Medicine Research Centers Program (<http://www.niaid.nih.gov>) through grant no. AI 39461 (GMD). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: tom.wilke@allzool.bio.uni-giessen.de

Introduction

Schistosomiasis japonica is a major parasitic disease threatening 50–65 million people living in subtropical areas of China [1]. Though overall prevalence and intensity of infection were reduced by more than 90% during the second half of the past century [2,3], the possibility of continued reduction of schistosomiasis to achieve rapid elimination has recently been questioned [4]. Highly variable rates of reduction across counties, continued persistence in some areas, and cases of re-emergence in others, remain major concerns [3,5,6]. The conditions that characterize the current, critical stage of disease elimination in China call for new strategies in disease surveillance and control [7]. The current control target aimed at reducing human and bovine infection rates in all endemic counties to less than 1% by 2015 [8–11] largely focuses on morbidity control. This strategy could benefit from the

inclusion of evolutionary and ecological perspectives, particularly as concerns key epidemiological and surveillance concepts.

For instance, a basic epidemiological concept used in China's schistosomiasis surveillance and control strategy is 'endemic area'. It refers to a region where a particular disease is prevalent [12] based on standardized parameters, mainly rates of infection in residents and/or cattle [13]. It does not explicitly consider evolutionary aspects such as the relative spatial isolation of populations transmitting the disease [7,14]. Such an evolutionary (i.e., population-based) approach, however, could help shift capabilities from simply analyzing the patterns of disease transmission to understanding the actual processes responsible for generating these patterns.

Another instance where an evolutionary perspective could be useful relates to the fact that China's current control strategy primarily targets only two main hosts, humans and cattle [15],

Author Summary

Schistosomiasis is considered the second most devastating parasitic disease after malaria. In China, it is transmitted to humans, cattle and other vertebrate hosts by a single intermediate snail host. It has long been suggested that the close co-evolutionary relationship between parasite and intermediate host makes the snail a major transmission bottleneck in the disease life cycle. Here, we use a novel approach to model the disease distribution in eastern China based on a combination of epidemiological, ecological, and genetic information. We found four major high risk areas for schistosomiasis occurrence in the large lakes and flood plain regions of the Yangtze River. These regions are interconnected, suggesting that the disease may be maintained in eastern China in part through the annual flooding of the Yangtze River, which drives snail transport and admixture of genotypes. The novel approach undertaken yielded improved prediction of schistosomiasis disease distribution in eastern China. Thus, it may also be of value for the predictive modeling of other host- or vector-borne diseases.

even though more than 40 mammalian species are known to serve as definitive host [7]. Therefore, it is reasonable to assume that cases of cryptic persistence of schistosomiasis are common, calling for approaches focusing on the snail host as major transmission bottleneck in the schistosome life cycle (“no snails, no disease”) [16].

Unlike other snail-schistosome models, *Schistosoma japonicum* is carried by a single species of snail with a peculiar amphibious life style, and there are no known cases of snail host-switching and/or host-addition in China [16]. This makes *Oncomelania hupensis* a crucial target for disease control. Empirical evidence even suggests a close link between snail genetic characters and rates of infection in snails [16–19], a possible consequence of the effects of positive frequency dependent selection, e.g., genetic warfare between host and parasite leading to sustained oscillations in genotype frequencies [20–22].

Given the large number of definitive hosts for *S. japonicum*, and the fact that quantitative parasite data are increasingly difficult to obtain due to low rates of natural infections [7], the intermediate snail host thus becomes of particular interest. Here, we take a fresh look at schistosomiasis distribution in eastern China from the evolutionary and ecological viewpoint of a single intermediate host system. For the first time, epidemiological with spatially explicit population-genetic and high-resolution ecological data are combined to develop a predictive model capable of estimating the probability that schistosomiasis occurs in a target area (here termed ‘spatially explicit schistosomiasis risk’—SESR). The study pursues four specific goals:

- We test whether genetic intermediate host characters (i.e., intrinsic evolutionary properties of populations), in principle, are reflected by previously defined endemic areas.
- We use a candidate set of topographical, ecological and genetic variables together with maximum entropy modeling to identify those explanatory characters that significantly discriminate between endemic/non-endemic areas.
- Based on the parameters with the highest discriminatory power, we then develop a SESR model, converting categorical infection data for administrative units into spatially explicit, high-resolution, and quantitative reaction data.

- Finally we compare the model data with actual human infection rates to evaluate our SESR model and to assess the significance of intermediate host traits for future epidemiological modeling of schistosomiasis.

Materials and Methods

Specimens Studied

Today, the human blood fluke *S. japonicum* (Katsurada, 1904) is transmitted in the Yangtze River area by two snail subspecies, *Oncomelania hupensis hupensis* Gredler, 1881 and *O. h. robertsoni* Bartsch, 1946. The two taxa (and therefore the disease as well) have disjunct ranges; the eastern subspecies *O. h. hupensis* occurs in the lowlands of the Yangtze River below the Three Gorges (Fig. 1); the western subspecies *O. h. robertsoni* in mountainous regions of Sichuan and Yunnan provinces.

This study includes 530 specimens of the eastern subspecies *O. h. hupensis* from 45 sites (‘populations’) in six Chinese provinces. Specimens were largely collected between 1996 and 2005 (plus three additional populations in 1984) in and around the lower Yangtze River basin. Thirty one of these sites are located within and fourteen outside of previously proposed endemic areas (Fig. 1; Supporting Table S1).

All specimens were obtained before the completion of the Three Gorges Dam (TGD) project in 2007. It is assumed that dam-associated changes in water regime and sedimentation rates of the lower Yangtze River will have significant effects on the distribution patterns of the intermediate host and thus disease distribution [10,16,23]. As these demographic processes are likely reflected by snail population structures [16], the data presented here may also serve as valuable baseline for future studies of pre- vs. post-dam effects on schistosomiasis in China.

Molecular Data

DNA extraction, amplification and sequencing. Genomic DNA was extracted from individual snail specimens utilizing a CTAB protocol [24]. Digital images of selected specimens were taken prior to consumptive DNA isolation and deposited at the University of Giessen Systematics and Biodiversity collection (UGSB). We amplified a fragment of the mitochondrial cytochrome *c* oxidase subunit I (COI) gene with a target length of 658 base pairs (excluding primer sequence). Forward and reverse primers for PCR amplification and DNA sequencing were LCO1490 [25] and COR722b [26]; the latter is a modification of primer HCO2198 [26]. Bidirectional DNA sequencing according to the ‘Sanger’ chain-termination method [27] was performed either on a Long Read IR2 4200 sequencer (LI-COR, Lincoln, NE, USA) or an ABI3730XL sequencer (Life Technologies Corporation, Carlsbad, CA, USA). The protein-coding COI sequences, which are free of insertions and deletions in the family Pomatiopsidae [28], were aligned in BioEdit 7.1.3.0 [29]. As the first base pairs behind the 3′ end of each primer were difficult to read, we trimmed these regions, leaving a 638 bp-long overlapping fragment. All 440 newly generated sequences were deposited in GenBank. Additional 90 sequences were taken from GenBank, resulting in a total dataset of 530 sequences (Supporting Table S1).

Correlation Analysis of Previously Defined Endemic Areas and Intermediate Host Genetics

In order to test whether genetic snail parameters, in principle, can be explained by the distribution of previously defined endemic areas (see Goal 1), we used two independent approaches. With an Analysis of Molecular Variance (Amova) [30], it was tested

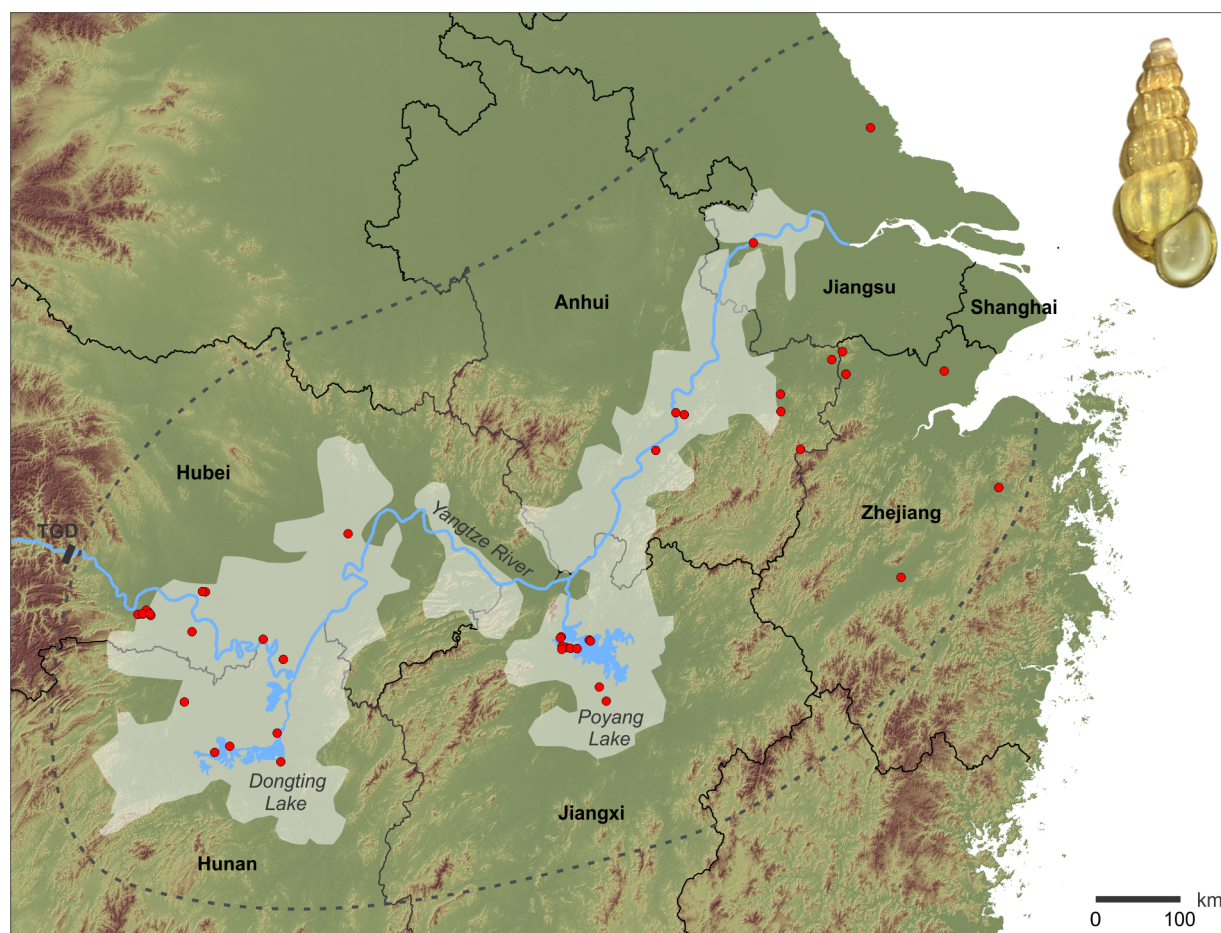


Figure 1. Schistosomiasis study area in eastern China. The map shows the localities of the intermediate snail host *Oncomelania h. hupensis* sampled (red dots), the assumed maximum distribution area of this subspecies in the lower Yangtze River basin (dashed gray line), and previously delineated endemic areas [5] (highlighted areas). The distribution area is based on our own sampling data and literature records [58,74,80,81], restricted by a reasonable vertical distribution of 0 to 200 m a.s.l. [2]. For detailed locality information see Supporting Table S1. TGD = Three Gorges Dam.

doi:10.1371/journal.pntd.0002327.g001

whether there is a significant partitioning of variance of genetic characters from populations within vs. populations outside of endemic areas. This analysis is not fully spatially-explicit as the overall geographical distribution of snail populations is not considered. Alternatively, we used Multivariate Regression Trees (MRT) [31] to test whether genetic characters can be predicted using endemic areas. This analysis differs from the Amova by being spatially-explicit (see below).

As our snail populations were collected over a period of approximately 20 years (see Supporting Table S1), we conducted two Amova tests (A, B). Amova A served as a pre-analysis to exclude the possibility of a sampling bias by testing for the presence of a significant partitioning of variance of genetic characters from populations being collected during different time periods. Amova B was then used to test for a partitioning of variance between populations from endemic vs. non-endemic areas. For the former analysis, our grouping variables were sampling periods (1980s, 1990s, and 2000s) according to Supporting Table S1. For the latter one, populations were grouped according to their assignment to endemic vs. non-endemic areas based on the endemic area distribution as previously suggested [5] (also see Fig. 1 and Supporting Table S1). Then, a distance matrix of pairwise nucleotide differences was

calculated in Arlequin 3.5.1.2. [32] and the significance of the Φ statistic ($\alpha \leq 0.05$) tested by generating a null-distribution based on 10,000 permutations of the original dataset.

The MRT approach used the principle coordinates obtained by multidimensional scaling of pairwise differences as response variables. The variables were standardized by dividing by the respective maximum, and endemic state and geographical coordinates served as explanatory variables (latitude and longitude information was converted into northing and easting according to the Asia North Equidistant Conic projection for obtaining equidistant values). As previously suggested [33], the first split of the tree can be forced for hypothesis testing. Accordingly, we pre-defined this split to discriminate between endemic and non-endemic areas. Hierarchical nesting for MRT was then done with the MVPARTwrap 0.1.8 package [33] for R 2.15 statistical environment [34]. The overall best tree was selected by running 1000 10-fold cross validations.

SESR Modeling

Candidate snail traits. For identifying the genetic parameters that significantly reflect the spatial distribution of endemic areas, four candidate population indices were calculated from the COI dataset. They comprised within-site ("diversity") and between-

site ('divergence') parameters. The first set of parameters consisted of nucleotide diversity π (average number of nucleotide differences per site within populations based on equation 10.5 in Nei [35]) and Tajima-Nei-distance D_{TN} (average number of nucleotide differences per site between populations, corrected for unequal rates of substitution [36]), calculated in the R statistical environment. The second set of indices utilized haplotype information, which was previously suggested to be correlated with infection rates in snails [2,16,19] (also see section "Predictive variables of SESR" in the Discussion). We calculated both haplotype diversity (H_D) and haplotype divergence (H_{MH}) in order to assess within and between population differentiations, respectively. H_D was estimated in DnaSP v5 based on equation 8.4 in Nei [35]. For calculating H_{MH} , we treated haplotypes as species [37] and estimated the dissimilarity between the haplotype structures of two groups in R 2.15 applying the Morisita-Horn index [38]. Note that the two divergence indices D_{TN} and H_{MH} were obtained by estimating the average pairwise distances between the population of concern and all other populations in the dataset.

Additional candidate ecological and topographical parameters of *O. hupensis* to be used for the modeling of schistosomiasis risks were selected based on their potential relevance for the distribution and/or susceptibility of snail populations [16,39–41] (see Table 1).

Model building. In order to identify the candidate characters that discriminate between endemic/non-endemic areas and to build and visualize our SESR model, we here used maximum entropy modeling as implemented in MaxEnt 3.3.3k [42]. The software has been shown to perform well in species distribution [43] and disease modeling analyses [44]. The MaxEnt algorithm fits predictor variables (i.e., our snail traits) to the endemic state by discriminating between the 31 populations located within previously defined endemic areas [5] and 10,000 random extralimital points. Our genetic data were based on 45 distinct collection points and hence did not have the continuity of the topographical and ecological characters. We therefore first interpolated the gaps between populations using Inverse Distance Weighting (IDW) [45] as implemented in the gstat 1.0–14 package [46] for the R

statistical environment. The IDW algorithm estimated values for the population indices by weighting the information of the nearest twelve sampling points with distance to the respective grid cell.

As correlated variables may lead to a decrease of model quality [47], we first tested similar variables (the four genetic variables, bio06 vs. bio11, and bio12 vs. bio16) in R for potential correlation using a conservative Pearson's r of 0.8 *sensu* Elith et al. and Rodda et al. [48,49] as threshold. Whereas no correlation could be detected among the genetic variables, variable pairs bio6/bio11 and bio12/bio16 were correlated with $r = 0.89$ and $r = 0.85$, respectively. The final selection of the best combination of topographical, ecological and genetic candidate variables was done in MaxEnt by determining the area under the receiver operator curve (AUC), with increasing numbers of variables being penalized. In order to account for the presence of two pairs of correlated variables, we ran four individual MaxEnt analyses (all uncorrelated variables+bio06+bio12; all uncorrelated variables+bio06+bio16; all uncorrelated variables+bio11+bio12; and all uncorrelated variables+bio11+bio16). For each variable combination, 5-fold cross validation was done with 50 repeats. Individual variable contributions were evaluated by jackknife testing of significant differences in AUC values, applying parametric bootstrapping with 10,000 replicates. The best combination of variables that fulfilled the quality criteria was then used for the final run of 500 predictions. In order to avoid false positive predictions [50], we assessed the risk threshold for each prediction using a receiver operator curve plot (ROC) [51] with predictions below the threshold being omitted. After calculating the mean of the 500 final run predictions, the information was processed with the GIS-package dismo 0.7–23 [52] in R to visualize our SESR model.

Linking Model Data for Schistosomiasis Risk with Infection Data

To evaluate our model and to assess its epidemiological value, we compared model predictions to observed human infection rates at county levels. Total numbers of human cases for each county were derived from Zhang et al. [53] and are based on

Table 1. Candidate topographical and ecological characters of *Oncomelania h. hupensis* used for the SESR modeling.

Parameter	Source/data transformation	Original resolution	Relevance
Elevation	SRTM3 90 m Digital Elevation Model	90 m	Main snail distribution parameter
Slope	SRTM3 90 m Digital Elevation Model	90 m	Main snail distribution parameter
Bioclimatic variable bio6 (minimum temperature of coldest month)	Global Climate database at www.worldclim.org [82]	1000 m	Lethal temperature for <i>O. hupensis</i> is -2.7°C [71]
Bioclimatic variable bio11 (mean temperature of coldest quarter)	Global Climate database at www.worldclim.org [82]	1000 m	The development of both snails and parasite larvae requires a minimum temperature [72,83,84]
Bioclimatic variable bio12 (annual precipitation)	Global Climate database at www.worldclim.org [82]	1000 m	Proxy for suitable snail habitat [84]
Bioclimatic variable bio16 (precipitation of wettest quarter)	Global Climate database at www.worldclim.org [82]	1000 m	Proxy for flooding, transporting and/or potentially drowning of snails
Euclidean distances to water bodies	Calculated in ArcMap 9.3 based on water body data in www.diva-gis.org	90 m	Proxy for suitable snail habitat and/or flooding
Normalized Difference Vegetation Index (NDVI)	Moderate Resolution Imaging Spectroradiometer (MODIS) data for 2000–2010, United States Geological Survey. Clouds were masked and the ten year average was calculated by using the raster 2.0–12 package [85] for R 2.15	250 m	Proxy for soil moisture [8] and therefore suitable snail habitat (wetlands)

All original resolutions were re-sampled to 500 m.
doi:10.1371/journal.pntd.0002327.t001

epidemiological studies conducted in 1999–2001, 2007 and 2008. Average human infection rates were then expressed as average ratio of total number of incidences and population sizes for the respective area and year. Year-specific human population sizes were obtained from the NASA Socioeconomic Data and Applications Center (SEDAC) available at <http://sedac.ciesin.columbia.edu/gpw>.

We note that this validation routine is not completely independent: the state of endemic areas used as response variable is based on local environment and prevalence of infection [5], with the latter being derived from cattle and/or human [13]. However, as we only used non-human and non-cattle traits for our subsequent risk modeling, a validation of this model by actual infection data can be informative.

We tested the infection dataset for spatial autocorrelation utilizing Moran's *I* Test [54]. Semiparametric Eigenvector filtering [55], as implemented in the R-package *spdep* 0.5–51 [56], was performed to quantify the potential effect of spatial autocorrelation. Then a linear regression analysis was conducted in R by linking the county level infection rates to the mean model prediction values for the respective areas. The *anova* function in R was used to assess the contribution of the explanatory variables relative to spatial eigenvectors and to correct the values accordingly.

Results

Genetic Indices

Genetic indices for individual populations are provided in Supporting Table S2. Our dataset of 530 specimens consisted of 212 haplotypes. Values for haplotype diversity (H_D) and haplotype divergence (H_{MH}) ranged from 0.0 to 1.0 and 0.801 to 1.0, respectively. Nucleotide diversity (π) and Tajima-Nei-distance (D_{TN}) varied from 0.0 to 0.016 and from 0.0 to 0.019, respectively. Spatially explicit heat maps of genetic values for indices that were later used for the SESR modeling (i.e., D_{TN} and H_D) are given in Supporting Figure S1.

Correlation Analysis of Previously Defined Endemic Areas and Snail Genetics

Analysis of molecular variance (Amova). The Φ_{CT} value for individuals collected during different sampling periods (i.e., 1980s, 1990s, and 2000s; Amova A) was -0.016 ($p = 0.464$). As negative values should be interpreted as zero [57] and as the high p -value indicates that our hypothesis (i.e., that the means of the groups are equal) is not rejected, a possible bias in our SESR modeling caused by different sampling periods appears to be unlikely.

The variation among groups of populations collected in endemic vs. non-endemic areas (Amova B) explained 18.1% of the total variation ($\Phi_{CT} = 0.18$, $p < 0.001$), the variation among populations within groups explained 40.4% ($\Phi_{SC} = 0.49$, $p < 0.001$), and the variation within populations 41.6% ($\Phi_{ST} = 0.58$, $p < 0.001$). As the Φ_{CT} value is relatively high and as there is a significant partitioning of variance of snail parameters from populations belonging to endemic areas vs. extralimital populations, genetic characters appear to explain in part the endemic area state (see Goal 1).

Multivariate Regression Trees (MRT). Overall, the spatially-explicit MRT explained 62.79% of the total variance in the dataset. The pre-defined first split between endemic and non-endemic areas contributed a substantial 24.14% to this variance, confirming the results of the Amova B (see previous section).

SESR Modeling

Based on the four individual MaxEnt runs and a possible combinations of a total of twelve candidate variables, the combination of altitude, $bio11$, $bio12$, D_{TN} , and H_D was selected by MaxEnt (Goal 2; see Fig. 2).

The goodness-of-fit (AUC) of this model had a median value of 0.97 (95% confidence limit: 0.80–1.00). When considering the two genetic and the three topographical/environmental parameters alone, the goodness-of-fit was 0.84 (0.57–1.00) and 0.92 (0.68–1.00), respectively (Fig. 3). The average goodness-of-fit value for two of the three topographical/environmental parameters was 0.86 ($N = 3$; details not shown here) and thus very similar to the value of the two genetic parameters (0.84). Parametric bootstrapping showed significant differences in means between the model with all variables and the model without genetic parameters. The mean risk threshold as indicated by the ROC plot was 0.115.

The visualization of the SESR model (Goal 3), conducted with the R-package *dismo*, is shown in Fig. 4. Accordingly, the predicted schistosomiasis risk is highest in the regions of (i) Dongting Lake and (ii) Poyang Lake, as well as in the Yangtze River floodplains and islands in (iii) Hanyang and Jianli counties (Hubei Province), and in (iv) Tongling and Guichi counties (Anhui Province). Moderate to high risks were inferred for the remaining floodplains of the Yangtze River in Hubei and Anhui provinces.

Linking SESR Model Data with Human Infection Data

Our spatial risk model was compared to human infection rates [53] (Goal 4). Significant spatial autocorrelation was observed within a distance of 100 km using Moran's *I* test. The Anova conducted showed that linear regression including spatial eigenvectors significantly explained more variation than without ($r^2 = 0.476$ and 0.238 , respectively). Removing spatial autocorrelation from the infection dataset resulted in a final adjusted r^2 of 0.338 ($p < 0.001$, $N = 284$). Thus, SESR model predictions are correlated with human infection rates.

Discussion

The basic findings of our study were: (i) intermediate host genetic parameters inferred from the COI gene are correlated with the distribution of previously defined endemic areas (Goal 1); (ii) the maximum entropy modeling suggested five explanatory variables (altitude, $bio11$, $bio12$, D_{TN} , and H_D) to discriminate between endemic/non-endemic areas (Goal 2); (iii) the visualization of our SESR model indicated the highest risks for the regions of Dongting and Poyang lakes as well as some floodplains of the Yangtze River in Hubei and Anhui provinces (Goal 3); and (iv) our model predictions are correlated with human infection rates (Goal 4). These findings are discussed further below in relation to the spatial distribution of endemic areas, the quality of predictive intermediate host traits (particularly genetic traits), and the implications of our risk modeling for future schistosomiasis surveillance and control strategies.

High Risk Areas

Model predictions suggested four relatively distinct, yet not fully isolated, areas as high risk regions. These areas, and the lack of complete isolation thereof, fit relatively well the spatial distribution of schistosomiasis previously suggested based on Bayesian random-effect modeling of reported schistosomiasis cases [53]. The areas include the two major lake systems Dongting and Poyang, which are at the center of schistosomiasis control in eastern China [11]. They have long been considered to be endemic areas [58] and a

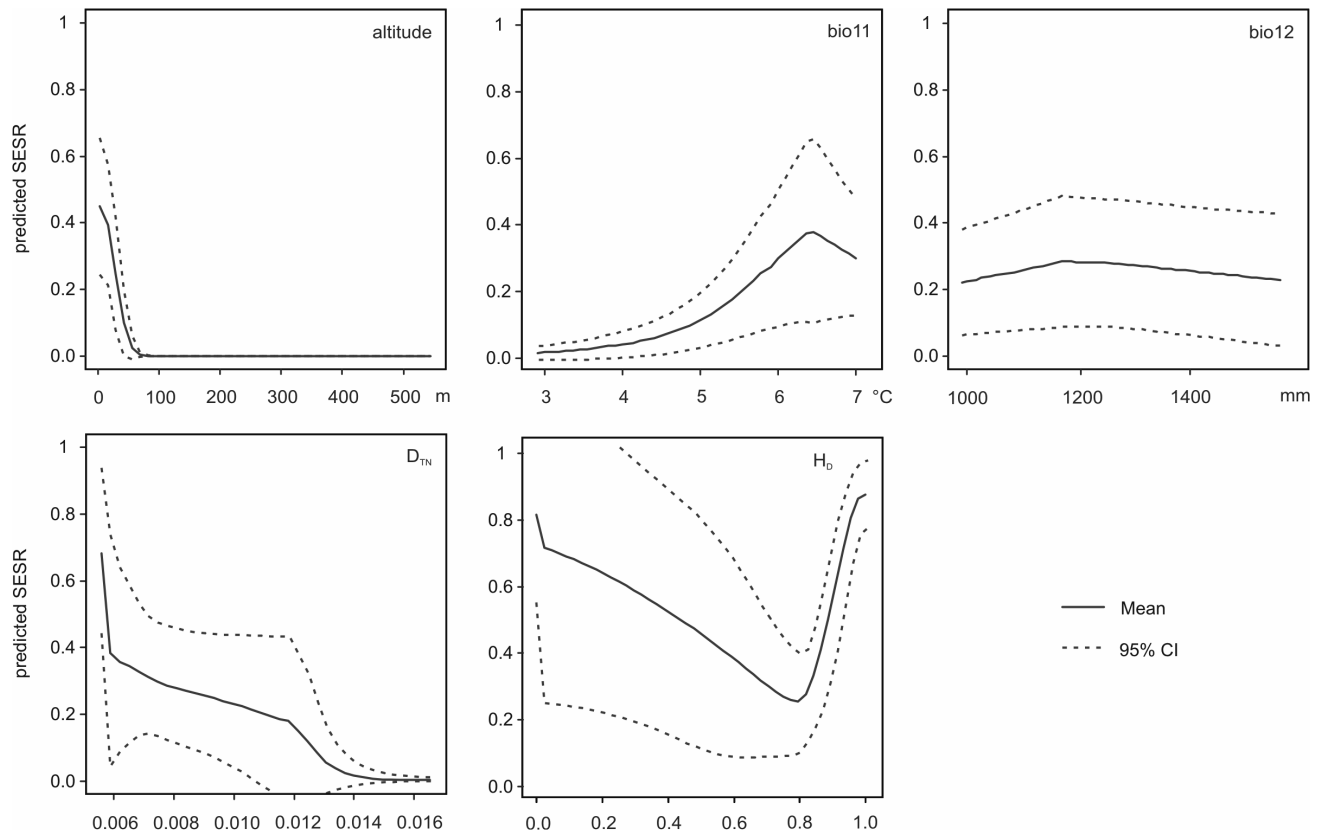


Figure 2. Individual response plots of five variables used for the SESR modeling. The plots were generated with the function response in the R-package dismo based on 500 model runs. Bio11 = mean temperature of coldest quarter, bio12 = annual precipitation, D_{TN} = Tajima-Nei-distance, H_D = haplotype diversity.
doi:10.1371/journal.pntd.0002327.g002

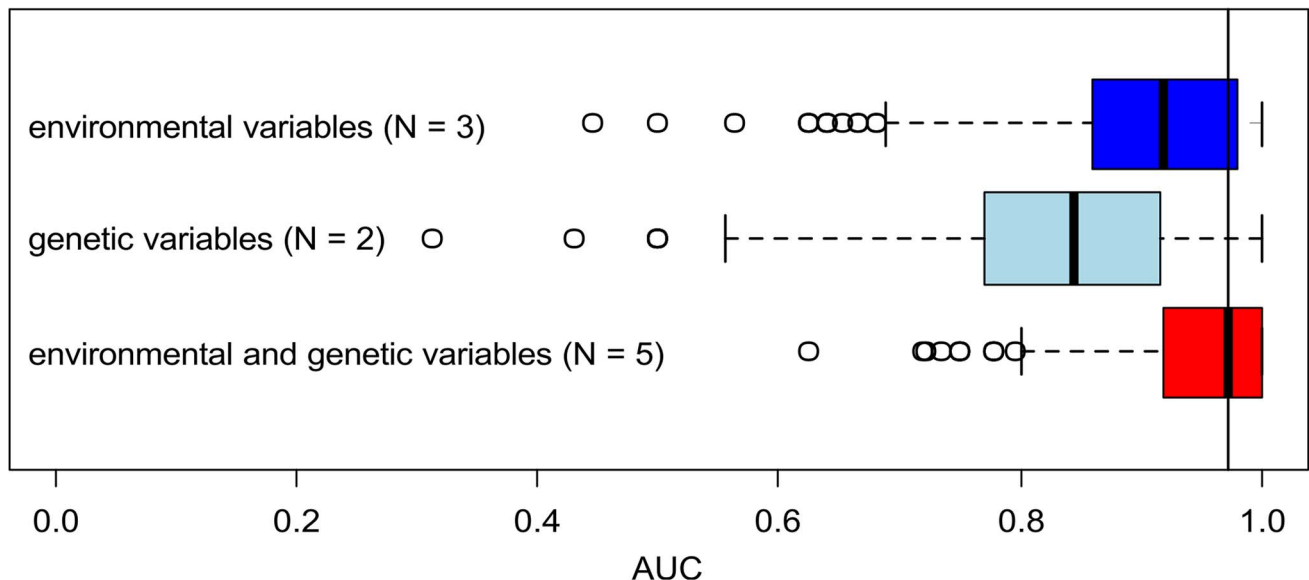


Figure 3. Results of jackknife testing of variable importance for the SESR modeling. The boxplots show the median goodness-of-fit values (AUC) of the models based on three environmental (bio11, bio12, altitude), two genetic (D_{TN} , H_D), and all five variables together with their respective 95% confidence limits (whiskers).
doi:10.1371/journal.pntd.0002327.g003

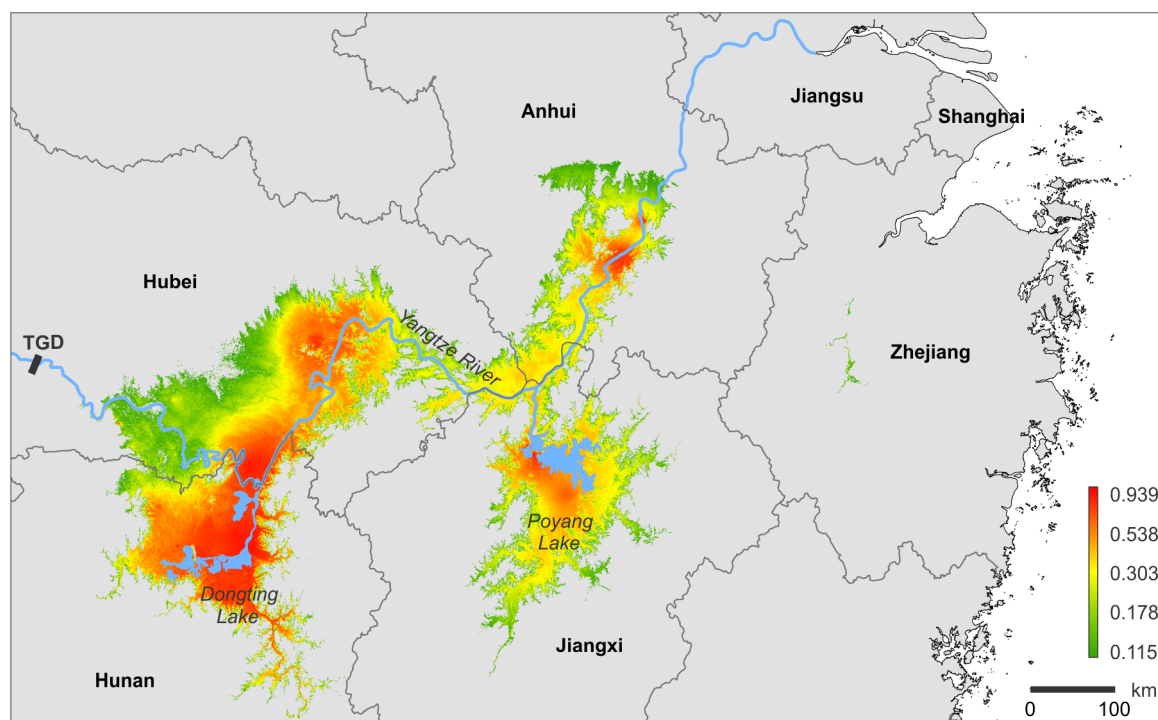


Figure 4. Output of the SESR modeling. Visualization of the schistosomiasis risk in eastern China (green color: low risk; red color: high risk). TGD = Three Gorges Dam.
doi:10.1371/journal.pntd.0002327.g004

high number of human re-infections occur there [13,59]. Both lake regions are heavily affected by the annual flooding of the Yangtze River and the associated rainy season (July–September) that causes a considerable increase of their surface areas [60,61]. These floods may have countervailing effects on the transmission of schistosomiasis in the lakes. Whereas they may result in large-scale drowning of adult snails and are therefore used in some regions as snail control measure [62], flooding and heavy rains generally promote schistosomiasis transmission in the region: (i) floods and associated sediment input create and sustain suitable snail habitat, (ii) floods are a major source of introduction and re-introduction of snails, (iii) floods lead to the admixture of different parasite lineages to which snail populations may not be well adapted, and (iv) inundation following heavy rains helps sustain suitable habitat for free-swimming parasite larvae.

Moreover, the lakes serve as major sediment traps for upstream Yangtze River sections during the flood season and as suppliers of suspended sediments for downstream river sections during the dry season [63–66]. In addition, the extensive floodplain areas of the lakes play an important role in flood control of downstream river sections. As a result, the parts of the Yangtze River upstream to the lakes are, generally, less suited as snail habitat than downstream parts. Our modeling confirmed these differential effects of the lakes on the Yangtze River. Whereas parts of the river upstream of the lakes only had low to medium risks, two high risk areas were located downstream of Lake Dongting (Hanyang and Jianli counties) and Lake Poyang (Tongling and Guichi counties). Overall, the enhancing effect of Lake Dongting seemed to be larger than that of Lake Poyang (Fig. 4). This could be explained with the upstream position of Lake Dongting. However, this could also partly result from a slight sampling bias (i.e., we studied comparable few snail populations from Yangtze River sections downstream of Lake Poyang). Interestingly, the high risk

regions inferred are not completely isolated. Therefore, re-introductions may play an important role for maintaining infections in populations [7] (also see section below).

In summary, the SESR modeling indicated four interconnected areas in the lower Yangtze River basin with high probabilities of disease occurrence. Risk values can vary considerably on small scales (i.e., within few kilometers) and are thus not associated with administrative entities. We suggest the complex hydrological interplay of lakes Dongting and Poyang with upstream and downstream sections of the Yangtze River in space and time as an important driver for the maintenance of the disease in eastern China.

Predictive Variables for SESR

Several studies have identified variables associated with schistosomiasis risk based on the life cycle of *S. japonicum* [2,39,40,67–70]. However, none of these studies utilized genetic characters, although their predictive value has long been suggested (see below). Our results confirmed the value of intermediate host genetic information for risk modeling. Two of the five final risk model variables were genetic characters (Fig. 2). In fact, the parameter with the single highest predictive value was haplotype diversity (H_D). In addition to this diversity parameter, the model also suggested a divergence index (D_{TN}) as risk variable, confirming the assumption that for understanding endemic areas and the potential isolation thereof, both genetic diversity (i.e., differences within populations) and divergence parameter (i.e., differences between populations) are of interest. In fact, one of the key findings of this study is that low divergence values indicate high risks (Fig. 2). In other words, not spatial isolation of endemic populations but high levels of gene flow and/or local effective population sizes drive schistosomiasis. This has been hypothesized before within the framework of frequency dependent selection

and/or in the context of demographic effects. The annual flooding of the Yangtze River, for example, may not only cause high levels of gene flow in snail but also in parasite populations, potentially leading to multiple infections or infection with parasites to which the snails are not adopted locally [16,19,24].

The remaining three variables—altitude, mean temperature of coldest quarter (bio11), and annual precipitation (bio12)—are all environmental (note that bio11 and bio12 are correlated with bio06 and bio16, respectively). These and several related variables have been previously suggested to be risk indicative [68,69,71]. In contrast, other environmental variables previously found to be important, such as water availability and vegetation index [68,72], did not significantly improve the goodness-of-fit of our model.

Overall, the relative contribution of the genetic parameters is very similar to the predictive power of the environmental ones. However, given the already high values of environmental and genetic parameters alone, complementing genetic with ecological variables and vice versa resulted in a significant but only slight increase in total goodness-of-fit. Nonetheless, genetic parameters appear to be particularly important as predictive variables as their application considerably reduces the confidence intervals of the predictions compared to environmental variables alone (i.e., by >40%; Fig. 3). As their addition helps improving the accuracy of the model, they are particularly helpful in areas with a high spatial risk dynamic. There, genetic parameters will likely enhance local risk modeling.

In summary, our results indicate that genetic intermediate host parameters can explain the endemic area state as well as the underlying evolutionary processes (i.e., ‘population’ isolation vs. admixture). The results also demonstrate the value of these parameters for risk modeling and for improving the local accuracy of model predictions. Both divergence and diversity parameters are of interest, and genetic and environmental parameters can be complementary in such an analysis.

Implications of SESR Modeling for Schistosomiasis Surveillance and Control Strategies

Our findings have several implications for schistosomiasis surveillance and control in China, particularly under the current situation of decreasing overall disease transmission rates and increasing potential for re-emergence:

- (i) With quantitative parasite data being extremely difficult to obtain, data of the co-evolved intermediate host can be used for risk modeling as previously suggested [16,18].
- (ii) Given that risk values can vary over small distances, fine-scale units should be used in place of administrative boundaries when making surveillance and control decisions.
- (iii) High-risk areas as derived from risk models should receive surveillance and control priority, yet given the complex hydrological interplay of water bodies in the Yangtze River floodplains, information from extralimital areas should be considered as well for local strategies.
- (iv) As population genetic parameters of the intermediate host significantly increase the quality of risk predictions, future routine genetic surveys of snail populations are encouraged, particularly within and near high risk areas to better understand both local population structures as well as regional patterns of population isolation and exchange among lakes and rivers.

In summary, this study stressed the role of intermediate host traits for understanding schistosomiasis occurrence in China in a

spatially-explicit manner. It also showed that these traits may serve as sensible proxies for infection risks and highlighted the potential of genetic characters for future risk modeling.

Limitations and Outlook

This study extended the traditional ecological niche-modeling approach, which is frequently used to predict the occurrence of parasite and/or host species, to an approach for predicting the probability of disease occurrence. This was possible by two key modifications/additions. First, we included epidemiological data (i.e., the spatial distribution of endemic areas) that were used to discriminate against. In a recent study on the West Nile virus mosquito vector [73], the authors demonstrated that such an approach was useful for predicting human incidences of West Nile virus. They also suggested that this method for creating probability distribution maps could be applied to the study of other vector-borne diseases. Second, we accounted for the problem that traditional niche-modeling approaches are typically based on (extrinsic) environmental data and not on intrinsic evolutionary information of the actual target populations. We here attempted to overcome this problem by including evolutionary and demographically relevant genetic information of the intermediate host in our disease occurrence modeling. To our best knowledge, this is the first study on vector-borne diseases that used this approach.

Whereas our findings could have important consequences for future schistosomiasis surveillance and control strategies, such study with pilot character also has some limitations. Given the considerable genetic diversity of *O. hupensis* in mainland China [24,74–76], our study of only 500+ specimens did likely not cover the full genetic structure of the intermediate snail host. Moreover, the mitochondrial COI gene used in this study can only reflect the phylogeographical and demographic history of populations; it is very likely not directly involved in co-evolutionary processes.

Therefore, we encourage future deep-sequencing based genomic and transcriptomic studies of the intermediate host that aim at identifying the genes responsible for susceptibility and/or resistance to infections (incl. horizontal gene transfer from parasite to snail). Furthermore, the view point of the intermediate host system as major transmission bottleneck has proven to be useful for understanding disease distribution. However, more specific evolutionary analyses such as identifying loci under selection (‘selective sweeps’) in high risk snail populations would help to better delineate co-evolutionary processes leading to rapid adaptation [77]. Using this information in future risk models will very likely further improve the predictive power of genetic information and may open new control perspectives. Finally, for an explicit modeling of actual infection risks in a multi-host parasite system, quantitative data on interspecies transmission dynamics are necessary [78,79]. This, in turn, would require future cross-disciplinary studies focusing on host-parasite ecological networks that consider all participating species [79].

Supporting Information

Figure S1 Spatially explicit heat maps of genetic indices used for the SESR modeling. Left: Tajima-Nei-distance (D_{TN}); right: haplotype diversity (H_D). (TIF)

Table S1 Locality information, number of specimens per site, and GenBank accession numbers for the total of 530 specimens of *Oncomelania h. hupensis* studied. Populations located inside of endemic areas are marked with an asterisk. (DOCX)

Table S2 Genetic population indices for 45 *Oncomelania h. hupensis* populations studied.
(DOCX)

Acknowledgments

The authors thank the employees of the Institute of Parasitic Diseases, Chinese National Center of Systematic Medical Malacology in Shanghai for their assistance with the collection of field samples. Silvia Nachtigall (Justus Liebig University) is gratefully acknowledged for her help with the

molecular work. Roland Schultheiß (Justus Liebig University) and three anonymous referees provided useful comments on a previous version of this paper.

Author Contributions

Conceived and designed the experiments: TW GMD JVR MS. Performed the experiments: TW MS TH ZZ. Analyzed the data: TW MS TH. Contributed reagents/materials/analysis tools: JMD FJ. Wrote the paper: TW MS.

References

- Gray DJ, Thrift AP, Williams GM, Zheng F, Li Y-S, et al. (2012) Five-year longitudinal assessment of the downstream impact on schistosomiasis transmission following closure of the Three Gorges Dam. *PLoS Negl Trop Dis* 6: e1588. doi:10.1371/journal.pntd.0001588.
- Ross AGP, Sleight AC, Li Y, Davis GM, Williams GM, et al. (2001) Schistosomiasis in the People's Republic of China: Prospects and challenges for the 21st century. *Clin Microbiol Rev* 14: 270–295. doi:10.1128/CMR.14.2.270-295.2001.
- Zhou X-N, Wang L-Y, Chen M-G, Wu X-H, Jiang Q-W, et al. (2005) The public health significance and control of schistosomiasis in China—then and now. *Acta Trop* 96: 97–105. doi:10.1016/j.actatropica.2005.07.005.
- Seto EYW, Remais JV, Carlton EJ, Wang S, Liang S, et al. (2011) Toward sustainable and comprehensive control of schistosomiasis in China: lessons from Sichuan. *PLoS Negl Trop Dis* 5: e1372. doi:10.1371/journal.pntd.0001372.
- Zhou X-N, Guo J-G, Wu X-H, Jiang Q-W, Zheng J, et al. (2007) Epidemiology of schistosomiasis in the People's Republic of China, 2004. *Emerg Infect Dis* 13: 1470–1476. doi:10.3201/cid1310.061423.
- Zhu R, Gray D, Thrift A, Williams GM, Zhang Y, et al. (2011) A 5-year longitudinal study of schistosomiasis transmission in Shian Village, the Anning River Valley, Sichuan Province, the People's Republic of China. *Parasite Vector* 4: 43. doi:10.1186/1756-3305-4-43.
- Carlton EJ, Bates MN, Zhong B, Seto EYW, Spear RC (2011) Evaluation of mammalian and intermediate host surveillance methods for detecting schistosomiasis reemergence in southwest China. *PLoS Negl Trop Dis* 5: e987. doi:10.1371/journal.pntd.0000987.
- Wang X-H, Zhou X-N, Vounatsou P, Chen Z, Utzinger J, et al. (2008) Bayesian spatio-temporal modeling of *Schistosoma japonicum* prevalence data in the absence of a diagnostic “gold” standard. *PLoS Negl Trop Dis* 2: e250. doi:10.1371/journal.pntd.0000250.
- Ellis MK, McManus DP (2009) Familial aggregation of human helminth infection in the Poyang Lake area of China with a focus on genetic susceptibility to schistosomiasis japonica and associated markers of disease. *Parasitology* 136: 699–712. doi:10.1017/S003118200900612X.
- McManus DP, Gray DJ, Li Y, Feng Z, Williams GM, et al. (2010) Schistosomiasis in the People's Republic of China: The era of the Three Gorges Dam. *Clin Microbiol Rev* 23: 442–466. doi:10.1128/CMR.00044-09.
- McManus DP, Gray DJ, Ross AGP, Williams GM, He H-B, et al. (2011) Schistosomiasis research in the Dongting Lake region and its impact on local and national treatment and control in China. *PLoS Negl Trop Dis* 5: e1053. doi:10.1371/journal.pntd.0001053.
- Cherath L, Frey R (2006) Leprosy. In: Longe JL, editor. *Gale encyclopedia of medicine*. Farmington Hills: Thomson Gale. pp. 2190–2193.
- Zhang S, Lin D (2002) The potential risk and control strategy in low endemic area of schistosomiasis in China. *Acta Trop* 82: 289–293. doi:10.1016/S0001-706X(02)00021-9.
- Miller FP, Vandome AF, McBrewster J (2010) *Endemic (epidemiology)*. International Book Marketing Service Limited. 104 p.
- Liang S, Spear RC, Seto EYW, Hubbard A, Qiu D (2005) A multi-group model of *Schistosoma japonicum* transmission dynamics and control: model calibration and control prediction. *Trop Med Int Health* 10: 263–278. doi:10.1111/j.1365-3156.2005.01386.x.
- Davis GM, Wilke T, Zhang Y, Xu X-J, Qiu C-P, et al. (1999) Snail-Schistosoma, Paragonimus interactions in China: Population ecology, genetic diversity, coevolution and emerging diseases. *Malacologia* 41: 355–377.
- Davis GM, Wilke T, Spolsky C, Qiu C-P, Oiu D-C, et al. (1998) Cytochrome oxidase I-based phylogenetic relationships among the Pomatiopsidae, Hydrobiidae, Rissoiidae and Truncatellidae (Gastropoda: Caenogastropoda: Rissoacea). *Malacologia* 40: 251–266.
- Davis GM, Wu W-P, Xu X-J (2005) Ecogenetics of shell sculpture in *Oncomelania* (Gastropoda) in canals of Hubei, China, and relevance for schistosome transmission. *Malacologia* 48: 253–264.
- Shi CH, Wilke T, Davis GM, Xia MY, Qiu CP (2002) Population genetics, micro-phylogeography, ecology, and susceptibility to schistosome infection of Chinese *Oncomelania hupensis hupensis* (Gastropoda: Rissooidea: Pomatiopsidae) in the Miao River system. *Malacologia* 44: 333–347.
- Bell G (1982) *The Masterpiece of nature: The evolution and genetics of sexuality*. CUP Archive. 650 p.
- Lively CM (2001) Trematode infection and the distribution and dynamics of parthenogenetic snail populations. *Parasitology* 123: 19–26. doi:10.1017/S0031182001008113.
- Webster JP, Davies CM (2001) Coevolution and compatibility in the snail-schistosome system. *Parasitology* 123: 41–56. doi:10.1017/S0031182001008071.
- Hotez PJ, Zheng F, Long-qi X, Ming-gang C, Shu-hua X, et al. (1997) Emerging and reemerging helminthiases and the public health of China. *Emerg Infect Dis* 3: 303–310.
- Wilke T, Davis GM, Qiu D, Spear RC (2006) Extreme mitochondrial sequence diversity in the intermediate schistosomiasis host *Oncomelania hupensis robertsoni*: Another case of ancestral polymorphism? *Malacologia* 48: 143.
- Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Marine Biol Biotechnol* 3: 294–299.
- Wilke T, Davis GM (2000) Intraspecific mitochondrial sequence diversity in *Hydrobia ulvae* and *Hydrobia ventrosa* (Hydrobiidae: Rissooidea: Gastropoda): Do their different life histories affect biogeographic patterns and gene flow? *Biol J Linn Soc* 70: 89–105. doi:10.1006/bjil.1999.0388.
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74: 5463–5467.
- Wilke T, Haase M, Hershler R, Liu H-P, Misof B, et al. (2013) Pushing short DNA fragments to the limit: Phylogenetic relationships of “hydrobioid” gastropods (Caenogastropoda: Rissooidea). *Mol Phylogenet Evol* 66: 715–736. doi:10.1016/j.ympev.2012.10.025.
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic acids symposium series* 41: pp. 95–98.
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479–491.
- De'ath G (2002) Multivariate regression trees: a new technique for modeling species-environment relationships. *Ecology* 83: 1105–1117. doi:10.2307/3071917.
- Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Res* 10: 564–567. doi:10.1111/j.1755-0998.2010.02847.x.
- Ouellette M-H (2012) MVPARTwrap: Additional functionalities for package mvpart. Available: <http://cran.r-project.org/web/packages/MVPARTwrap/index.html>.
- R Development Core Team (2012) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available: <http://www.R-project.org>.
- Nei M (1987) *Molecular evolutionary genetics*. New York: Columbia University Press. 526 p.
- Tajima F, Nei M (1984) Estimation of evolutionary distance between nucleotide sequences. *Mol Biol Evol* 1: 269–285.
- Helmus MR, Bland TJ, Williams CK, Ives AR (2007) Phylogenetic measures of biodiversity. *Am Nat* 169: E68–E83.
- Horn HS (1966) Measurement of “overlap” in comparative ecological studies. *Am Nat* 100: 419–424.
- Yang G-J, Vounatsou P, Zhou X-N, Tanner M, Utzinger J (2005) A Bayesian-based approach for spatio-temporal modeling of county level prevalence of *Schistosoma japonicum* infection in Jiangsu Province, China. *Int J Parasitol* 35: 155–162. doi:10.1016/j.ijpara.2004.11.002.
- Zhang Z, Carpenter TE, Chen Y, Clark AB, Lynn HS, et al. (2008) Identifying high-risk regions for schistosomiasis in Guichi, China: A spatial analysis. *Acta Trop* 107: 217–223. doi:10.1016/j.actatropica.2008.04.027.
- Zhang Z, Ong S, Peng W, Zhou Y, Zhuang J, et al. (2008) A model for the prediction of *Oncomelania hupensis* in the lake and marshland regions, China. *Parasitol Int* 57: 121–131. doi:10.1016/j.parint.2007.09.008.
- Phillips SJ, Dudik M, Schapire RE (2004) A maximum entropy approach to species distribution modeling. *Proceedings of the twenty-first international conference on Machine learning*. ICML '04. New York, NY, USA: ACM. p. 83. Available: <http://doi.acm.org/10.1145/1015330.1015412>. Accessed 22 November 2012.
- Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, et al. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129–151. doi:10.1111/j.2006.0906-7590.04596.x.

44. Slater H, Michael E (2012) Predicting the current and future potential distributions of Lymphatic Filariasis in Africa using maximum entropy ecological niche modelling. *PLoS ONE* 7: e32202. doi:10.1371/journal.pone.0032202.
45. Shepard D (1968) A two-dimensional interpolation function for irregularly-spaced data. Proceedings of the 1968 23rd ACM national conference. ACM '68. New York, NY, USA: ACM. pp. 517–524. Available: <http://doi.acm.org/10.1145/800186.810616>. Accessed 30 October 2012.
46. Pebesma EJ (2004) Multivariable geostatistics in S: the gstat package. *Comput Geosci* 30: 683–691. doi:10.1016/j.cageo.2004.03.012.
47. Baldwin RA (2009) Use of maximum entropy modeling in wildlife research. *Entropy* 11: 854–866. doi:10.3390/e11040854.
48. Elith J, Kearney M, Phillips S (2010) The art of modeling range-shifting species. *Methods Ecol Evol* 1: 330–342.
49. Rodda GH, Jarnevich CS, Reed RN (2011) Challenges in Identifying Sites Climatically Matched to the Native Ranges of Animal Invaders. *PLoS ONE* 6: e14670. doi:10.1371/journal.pone.0014670.
50. Cantor SB, Sun CC, Tortolero-Luna G, Richards-Kortum R, Follen M (1999) A comparison of C/B ratios from studies using receiver operating characteristic curve analysis. *J Clin Epidemiol* 52: 885–892. doi:10.1016/S0895-4356(99)00075-X.
51. Bean WT, Stafford R, Brashares JS (2012) The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. *Ecography* 35: 250–258. doi:10.1111/j.1600-0587.2011.06545.x.
52. Hijmans RJ, Phillips SJ, Leathwick J, Elith J (2012) Dismo: species distribution modeling. Available: <http://cran.r-project.org/web/packages/dismo/index.html>.
53. Zhang Z, Zhu R, Ward MP, Xu W, Zhang L, et al. (2012) Long-term impact of the World Bank loan project for schistosomiasis control: A comparison of the spatial distribution of schistosomiasis risk in China. *PLoS Negl Trop Dis* 6: e1620. doi:10.1371/journal.pntd.0001620.
54. Moran PAP (1950) Notes on continuous stochastic phenomena. *Biometrika* 37: 17–23. doi:10.2307/2332142.
55. Tiefelsdorf M, Griffith DA (2007) Semiparametric filtering of spatial autocorrelation: the eigenvector approach. *Environ Plann A* 39: 1193–1221. doi:10.1068/a37378.
56. Bivand R (2012) Spdep: spatial dependence: weighting schemes, statistics and models. Available: <http://cran.r-project.org/web/packages/spdep/index.html>.
57. Lorenzo-Carballa MO, Hadrys H, Cordero-Rivera A, Andrés JA (2012) Population genetic structure of sexual and parthenogenetic damselflies inferred from mitochondrial and nuclear markers. *Heredity* 108: 386–395.
58. Li S-Z, Wang Y-X, Yang K, Liu Q, Wang Q, et al. (2009) Landscape genetics: the correlation of spatial and genetic distances of *Oncomelania hupensis*, the intermediate host snail of *Schistosoma japonicum* in mainland China. *Geospatial Health* 3: 221–231.
59. Guo J, Li Y, Gray D, Ning A, Hu G, et al. (2006) A drug-based intervention study on the importance of buffaloes for human *Schistosoma japonicum* infection around Poyang Lake, People's Republic of China. *Am J Trop Med Hyg* 74: 335–341.
60. Du Y, Cai S, Zhang X, Zhao Y (2001) Interpretation of the environmental change of Dongting Lake, middle reach of Yangtze River, China, by 210Pb measurement and satellite image analysis. *Geomorphology* 41: 171–181. doi:10.1016/S0169-555X(01)00114-3.
61. Shankman D, Keim BD, Song J (2006) Flood frequency in China's Poyang Lake region: trends and teleconnections. *Int J Climatol* 26: 1255–1266. doi:10.1002/joc.1307.
62. Davis GM, Zhang Y, Xu X, Yang X (1999) Allozyme analyses test the taxonomic relevance of ribbing in Chinese *Oncomelania* (Gastropoda: Rissoacea: Pomatiopsidae). *Malacologia* 41: 297–317.
63. Chen Z, Li J, Shen H, Zhanghua W (2001) Yangtze River of China: Historical analysis of discharge variability and sediment flux. *Geomorphology* 41: 77–91. doi:10.1016/S0169-555X(01)00106-4.
64. Li Y, Deng J, Sun Z, HE Y (2003) A study on the flood storage capacity in Dongting Lake. *Int J Sediment Res* 18: 138–147.
65. Dai S, Yang S, Zhu J, Gao A, Li P, et al. (2005) The role of Lake Dongting in regulating the sediment budget of the Yangtze River. *Hydrol Earth Syst Sc* 9: 692–698.
66. Chang J, Li J, Lu D, Zhu X, Lu C, et al. (2010) The hydrological effect between Jingjiang River and Dongting Lake during the initial period of Three Gorges project operation. *J Geogr Sci* 20: 771–786. doi:10.1007/s11442-010-0810-9.
67. Ross AGP, Bartley PB, Sleight AC, Olds GR, Li Y, et al. (2002) Schistosomiasis. *New Engl J Med* 346: 1212–1220. doi:10.1056/NEJMra012396.
68. Guo J-G, Vounatsou P, Cao C-L, Utzinger J, Zhu H-Q, et al. (2005) A geographic information and remote sensing based model for prediction of *Oncomelania hupensis* habitats in the Poyang Lake area, China. *Acta Trop* 96: 213–222. doi:10.1016/j.actatropica.2005.07.029.
69. Utzinger J (2005) Conquering schistosomiasis in China: the long march. *Acta Trop* 96: 69–96.
70. Yang J, Zhao Z, Li Y, Krewski D, Wen SW (2009) A multi-level analysis of risk factors for *Schistosoma japonicum* infection in China. *Int J Infect Dis* 13: e407–e412. doi:10.1016/j.ijid.2009.02.005.
71. Zhou X-N, Yang G-J, Yang K, Wang X-H, Hong Q-B, et al. (2008) Potential impact of climate change on schistosomiasis transmission in China. *Am J Trop Med Hyg* 78: 188–194.
72. Yang G-J, Vounatsou P, Zhou X-N, Tanner M, Utzinger J (2005) A potential impact of climate change and water resource development on the transmission of *Schistosoma japonicum* in China. *Parassitologia* 47: 127.
73. Larson SR, DeGroot JP, Bartholomay LC, Sugumaran R (2010) Ecological niche modeling of potential West Nile virus vector mosquito species in Iowa. *J Insect Sci* 10: 110.
74. Zhao QP, Jiang MS, Littlewood DTJ, Nie P (2010) Distinct genetic diversity of *Oncomelania hupensis*, intermediate host of *Schistosoma japonicum* in mainland China as revealed by ITS sequences. *PLoS Negl Trop Dis* 4: e611. doi:10.1371/journal.pntd.0000611.
75. Wilke T, Davis GM, Cui-E C, Xiao-Nung Z, Xiao Peng Z, et al. (2000) *Oncomelania hupensis* (Gastropoda: Rissooidea) in eastern China: Molecular phylogeny, population structure, and ecology. *Acta Trop* 77: 215–227. doi:10.1016/S0001-706X(00)00143-1.
76. Hauswald A-K, Remais JV, Xiao N, Davis GM, Lu D, et al. (2011) Stirred, not shaken: genetic structure of the intermediate snail host *Oncomelania hupensis robertsoni* in an historically endemic schistosomiasis area. *Parasite Vector* 4: 206. doi:10.1186/1756-3305-4-206.
77. Sabeti P (2008) Natural selection: Uncovering mechanisms of evolutionary adaptation to infectious disease. *Nature Education* 1. Available: <http://www.nature.com/scitable/topicpage/natural-selection-uncovering-mechanisms-of-evolutionary-adaptation-34539>. Accessed 17 December 2012.
78. Dobson A (2004) Population dynamics of pathogens with multiple host species. *Am Nat* 164: S64–S78. doi:10.1086/424681.
79. Rigaud T, Perrot-Minnot M-J, Brown MJF (2010) Parasite and host assemblages: embracing the reality will improve our knowledge of parasite transmission and virulence. *Proc R Soc B* 277: 3693–3702. doi:10.1098/rspb.2010.1163.
80. Kumagai T, Furushima-Shimogawara R, Ohmac H, Wang T-P, Lu S, et al. (2010) Detection of early and single infections of *Schistosoma japonicum* in the intermediate host snail, *Oncomelania hupensis*, by PCR and Loop-Mediated Isothermal Amplification (LAMP) assay. *Am J Trop Med and Hyg* 83: 542–548. doi:10.4269/ajtmh.2010.10-0016.
81. Yang G-J, Zhou X-N, Sun L-P, Wu F, Zhong B, et al. (2011) Compensatory density feedback of *Oncomelania hupensis* populations in two different environmental settings in China. *Parasite Vector* 4: 133. doi:10.1186/1756-3305-4-133.
82. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 25: 1965–1978. doi:10.1002/joc.1276.
83. Sun L, Zhou X, Hong Q, Huang Y, Yang G, et al. (2003) Initial temperature for the development of *Schistosoma japonicum* larvae in *Oncomelania hupensis* [in Chinese]. *Zhongguo Ji Sheng Chong Xue Yu Ji Sheng Chong Bing Za Zhi* 21: 303–306.
84. Yang G-J, Utzinger J, Sun L-P, Hong Q-B, Vounatsou P, et al. (2007) Effect of temperature on the development of *Schistosoma japonicum* within *Oncomelania hupensis*, and hibernation of *O. hupensis*. *Parasitol Res* 100: 695–700. doi:10.1007/s00436-006-0315-8.
85. Hijmans RJ, Van Etten J (2012) Raster: geographic analysis and modeling with raster data. Available: <http://cran.r-project.org/web/packages/raster/index.html>.